Rishabh Sanjay

B.S in Mathematics and Scientific Computing, Minor in ML and Applications

Indian Institute of Technology, Kanpur

Linkedin : rishabh-sanjay

Email : rishabh.lfs@gmail.com

Phone : +91-8825236258

Github : Rishabh261998

Academic Qualifications

Year	Degree/Certificate	Institute	GPA/%
2017 - 2021	B.S	Indian Institute of Technology, Kanpur	8.5/10
2017	ISC(XII)	Little Flower School, Jamshedpur	94.25%
2015	ICSE(X)	Little Flower School, Jamshedpur	95.60%

Publications

- **ILDC for CJPE: Indian Legal Documents Corpus for Court Judgment Prediction and Explanation** *[paper]* *[code]*
  V Malik, **R Sanjay**, SK Nigam, K Ghosh, SK Guha, A Bhattacharya, A Modi
  Association for Computational Linguistics and the International Joint Conference on NLP (ACL-IJCNLP)
- **Semantic Segmentation of Legal Documents via Rhetorical Roles** *[paper]* *[code]*
  V Malik*, **R Sanjay***, SK Guha, SK Nigam, A Hazarika, A Bhattacharya, A Modi (*Equal Contribution)
  Natural Legal Language Processing(NLLP) 2022 Workshop EMNLP

Technical Skills

- **Interests:** Machine Learning, Deep Learning, Natural Language Processing, Competitive Programming
- **Programming Languages:** C, C++, LaTeX, MATLAB, Python, Java, HTML, JS, SQL
- **Software and Libraries:** Pytorch, Keras, Scikit-learn, Numpy, Pandas, Matplotlib, Bokeh, Pytest, Git

Research Experience

- **WikiHow QA Dataset Project** *(May'22-Present)*
  Research Intern : Ohio State University Mentor: Prof. Sun Huan
    - Analysed a WikiHow FAQ QA dataset for the Cooking and DIY category.
    - Experimented on downstream tasks to check how **pre-training LM** on the custom dataset before **fine-tuning** on task.
    - Used **RoBERTa-large** for Question Type Classification, Re-Ranking, MRC and Intent Recognition tasks.
    - Used **T5-large** and **bloom** LLMs for **Abstractive Closed Book-ODQA** on the WikiHow FAQ dataset.
- **Court Judgement Prediction, Explanation on Indian Supreme Court Cases** *(Mar'20 - June'20)*
  Mentor: Prof. Ashutosh Modi and Prof. Arnab Bhattacharya
    - Created a corpus of **32,000** legal jugdements and tackled the open research problem of Court **Judgement Prediction**.
    - Achieved the best acc. of **78%** (using XLNet) after extensive experiments on 14 SOTA document classification methods.
    - Used different explainability methods such as **IG, Occlusion, Attention etc.** to address explainability.
    - Evaluated our explainability results using gold annotations by legal experts using **ROUGE, BLEU, METEOR etc.**
- **Semantic Segmentation of Legal Documents via Rhetorical Roles** *(April'21 - Aug'21)*
  Mentor: Prof. Ashutosh Modi and Prof. Arnab Bhattacharya
    - Created a new corpus of **100** legal cases, annotated with rhetorical role labels. This is the largest RR corpus.
    - We propose new multi-task learning model **MTL-BiLSTM-CRF(BERT-SC)** which uses label shift as an auxillary task.
    - Our model has an F1 score of **0.71** which is better than various other baselines and sequence classification models.
    - We also show an application of RR for the **judgement prediction** task.
- **SemEval2021: HaHackathon: Detecting and Rating Humor and Offense** *(Sep'20 - Dec'20)*
  Course Project: CS771, Mentor: Piyush Rai
    - Developed and trained various machine learning models for Humor classification.
    - Used various ML models from classical models to Transformer based models like **BERT+BiGRU** and achieved F1 of **0.94**.
    - Also developed a **BERT+MLP** model to give real number rating to the short text in the range 0-4, achieved a RMSE score of **0.446**.

Work Experience

- **Oracle** *(July'21 - Present)*
  Applications Engineer, Fusion HCM Development *(Hyderabad, IN)*
    - Developed new features for the Oracle Fusion HRMS tool using **Oracle ADF(Java-based Framework)**.
    - Worked on **RESTful** web services for Oracle Journeys.
    - Delivered various **enhancements** and **bug fixes** for both UI and back-end.
    - Analysed and resolved real-time **customer issues** related to the product.
- **Goldman Sachs** *(May'20 - June'20)*
  Summer Analyst, IMD Core Engineering Division *(Bangalore, IN)*
    - Created a **Horizontally Scalable** and **Distributed Reconciliation** System in Java for trading data.
    - Used the **divide and conquer methodology** and Akka library in Java to build the distributed system.

- – Created multiple **events** and distributed data so that the data can be processed **concurrently** across the systems.
  - – Queried **Mongo** and **Sybase IQ** databases for the reconciliation process and loaded any inconsistencies.
- **Google Summer of Code at NumFOCUS** *(May'21 - Aug'21)*
  Open-source Developer, ArviZ *(Remote)*
  - – Implemented plots for Bayesian Visualization like **Dot plots**, **ECDF plots** etc. with backend as **Matplotlib** and **Bokeh**
  - – Implemented various algorithms which were incorporated in the implementation of the above plots
  - – Used **pytest** library for writing tests and also added **examples** and **documentation** for easier understanding

Relevant Projects

- **Neural Machine Translation: Hindi to English** *(Jan'21- Mar'21)*
  (Course Project: CS779, IIT Kanpur)
  - – Implemented a 2 layered bi-directional **GRU** as Encoder and single layered as decoder with **attention** mechanism.
  - – Optimized convergence of model using Teacher forcing, and used the **Beam Search** strategy for decoding.
- **Mining and Analysing Indian Supreme Court and High Court Judgements** *(Sep'20- Dec'20)*
  (Course Project: CS685, IIT Kanpur)
  - – Scraped **3 lakh** cases and extracted relevant sections and clustered cases using **k-means** by IPC topics.
  - – Incorporated demographic factors of area and population to compare crime rate under different High Courts.
- **Breaking Cryptosystems** *(Jan'20- April'20)*
  (Course Project: CS641, IIT Kanpur)
  - – Broke **Substitution, Substitution-Permutation, DES, SASAS** and weaker variants of **RSA** and **KECCAK** ciphers.
- **Taxi Fare Forecasting** ⌂ code *(Oct'19- Nov'19)*
  (Course Project: Time Series Analysis, IIT Kanpur)
  - – Visualized **time-series data**, eliminated trends and fitted **ARIMA model** on the data for forecasting taxi fares.
- **Low Rank Matrix Approximations and Algorithms** ⌂ code *(May'19 - June'19)*
  (Mentor: Sumit Ganguly, IIT Kanpur)
  - – Implemented **sampling algorithms** for **matrix approximations** and **low rank approximation** of matrix.
  - – Implemented **length squared sampling based** matrix multiplication and **CUR method** for matrix sketching.
- **Reinforcement Learning** ⌂ code *(Dec'18)*
  - – Trained Gym's frozen-lake game using **value-iteration** and Mountain-Car-v2 using **MDP and Q-learning** techniques.

Scholastic Achievements

- **G**raduated with **distinction** at Convocation'21 at IIT Kanpur.
- **KVPY Scholarship Awardee**(2017) by Indian Institute of Science and Govt. of India.
- Received **INSPIRE scholarship** by DST, Govt of India.
- Secured **All India Rank 2770** in **JEE Advanced 2017** among **220,000** candidates.
- Secured **All India Rank 1633** in **JEE Mains 2017** among the **1.2** million candidates.
- Qualified for **Indian National Maths Olympiad(2016)** with state rank **12th** at **Regional Maths Olympiad**.
- Rated 1800+ on Codeforces and a 5 star on Codechef.

Relevant Courses

Data Structure and Algorithms	Introduction to Machine Learning	Probability and Statistics
Probabilistic Modelling and Inference	Linear algebra	Algorithms II
Statistical Natural Language Processing	Data Mining	Machine Translation
Online Courses		
Sequence Models(Certificate)	Fundamentals of RL(Certificate)	Tensorflow for AI, ML, DL(Certificate)

Positions of Responsibility

- **Academic Mentor, Counselling Service Team**: *(2018-19)*
  - – Conducted remedial classes, one-to-one mentoring, doubt classes and designed study material for academically weak students.
- **Senior Executive Stamatics-Society**: *(2018-19)*
  - – Organized Mathematical talks and Mathematical competitions every week for Professors and students.
- **Secretary, Robotics Club**: *(2018-19)*
  - – Mentored freshers and organized various lectures and workshops for students.